



The
University
Of
Sheffield.

BMS 353

Bioinformatics for Biomedical Science

Module coordinator: Dr Marta Milo
Research Software Instructor: Dr Mike Croucher

An Introduction to the tools we'll be using



R is a free language and environment for statistical computing and graphics.

Job Trends from Indeed.com

— R !"R D" !"A R" !"H R" !"R N" !toys !kids !" R Walgreen" !walmart !"HVAC R" !"R Bard" and (

— SPSS and ("big data" or "statistical analysis" or "data mining" or "data analytics" or "machir



How popular is R? <http://r4stats.com/articles/popularity/>

An Introduction to the tools we'll be using



Packages

Save time by using other people's code.

Comprehensive R Archive Network (CRAN)

7468 packages (November 2015)

<https://cran.r-project.org/web/packages/>



Bioconductor

1104 packages (November 2015)

<https://www.bioconductor.org/>

An Introduction to the tools we'll be using



Packages

Save time by using other people's code.

Comprehensive R Archive Network (CRAN)

9262 packages (October 2016)

<https://cran.r-project.org/web/packages/>



Bioconductor

1211 packages (October 2016)

<https://www.bioconductor.org/>

An Introduction to the tools we'll be using



Open source, interactive data science and scientific computing across over 40 programming languages.

An Introduction to the tools we'll be using

Currently in use at

Google

 Microsoft

IBM

Bloomberg

O'REILLY®

CONTINUUM
ANALYTICS

 **rackspace.**
the #1 managed cloud company


Quantopian

 NetApp®


software
carpentry

hhmi **janelia**
Research Campus

< CODE **NEURO** >

N-Site LLC


 SageMathCloud

BRYN
MAWR
COLLEGE

CAL POLY
SAN LUIS OBISPO

Berkeley
UNIVERSITY OF CALIFORNIA

 The
University
Of
Sheffield.

THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC


NORTHWESTERN
UNIVERSITY

 NYU



An Introduction to the tools we'll be using



Combine live computer code, data, text and mathematics in one interactive document.

Academic papers only give a description of your analysis.

Jupyter notebooks contain both the description and the analysis itself.



An Introduction to the tools we'll be using

jupyter Untitled1 Last Checkpoint: 20 hours ago (unsaved changes)

File Edit View Insert Cell Kernel Help

Cell Toolbar: None

```
In [22]: x <- rnorm(10)
y <- rnorm(10)
summary(lm(y~x))
```

Out[22]:

```
Call:
lm(formula = y ~ x)

Residuals:
    Min       1Q   Median       3Q      Max
-1.4317 -0.8257 -0.1580  0.9709  1.4319

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.01308    0.34428   0.038   0.9706
x            1.05533    0.44650   2.364   0.0457 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.088 on 8 degrees of freedom
Multiple R-squared:  0.4112,    Adjusted R-squared:  0.3376 
F-statistic: 5.586 on 1 and 8 DF,  p-value: 0.0457
```


An Introduction to the tools we'll be using



Collaborative computational
mathematics

An open source environment for running Jupyter
notebooks (and others!) in the cloud.

Dedicated Virtual Machine for this course on the
Google Cloud Platform.

As powerful as a node on the Sheffield University
Supercomputer: Iceberg

An Introduction to the tools we'll be using



All you need is a web browser and internet connection.



The
University
Of
Sheffield.

Log into SageMathCloud

<https://cloud.sagemath.com/>

Create account (or [sign in](#))

☐ First, agree to the [Terms of Service](#)

Use your email address

Name

First and Last Name

Email

Email

Choose a password

Choose a password

Create account for free

Or use

Google

Facebook

Github

Twitter

Not working? Email us at help@sagemath.com immediately!